CLAIMS

What is claimed is:

1. A system that facilitates spam detection comprising:

a component that receives an item and extracts a set of features associated with an origination of a message or part thereof and/or information that enables an intended recipient to contact or respond to the message; and

a component that analyzes a subset of the extracted features in connection with building and employing a plurality of feature-specific filters that are independently trained to mitigate undue influence of at least one feature type over another in the message, the subset of extracted features comprising of at least one of a URL and an IP address, and the plurality of filters comprising at least a first feature-specific filter.

2. The system of claim 1, further comprising a plurality of training components that individually employ at least one of IP addresses or URLs and other features, respectively, in connection with building the plurality of feature-specific filters.

3. The system of claim 1, the first feature-specific filter is trained using IP addresses.

4. The system of claim 1, the first feature-specific filter is trained using URLs.

5. The system of claim 1, the plurality of feature specific filters comprising a second feature-specific filter that is trained using a subset of features extracted from the message other than a URL and an IP address.

6. A system that facilitates spam detection comprising:

a component that receives an item and extracts a set of features associated with an origination of a message or part thereof and/or information that enables an intended recipient to contact or respond to the message;

at least one filter that is used when one of the IP address of the message or at least some part of at least one of the URLs in the message is unknown.

7. The system of claim 6, the at least one filter is trained using some number of bits less than 32 bits of an IP address.

8. The system of claim 1, further comprising a filter combining component that combines information collected from the first and second feature-specific filters.

9. The system of claim 8, the first feature-specific filter detects at least one of known IP addresses and at least one known URL in the message.

10. The system of claim 8, the second feature-specific filter detects non-IP address and non-URL data in the message.

11. The system of claim 8, the filter combining component combines the information by at least one of multiplying scores generated by the filters, adding scores generated by the filters, or training an additional filter to combine the scores.

12. The system of claim 6, the at least one filter is trained using all bits of an IP address.

13. The system of claim 6, further comprising a filter selection component that selects and employs at least one feature-specific filter out of the plurality of filters for which there is sufficient data extracted from the message.

14. The system of claim 1, the first feature-specific filter is trained independently of the second feature-specific filter to mitigate either filter influencing the other when filtering the message.

15. The system of claim 14, at least one of the feature-specific filters models dependencies.

16. The system of claim 1, the plurality of feature-specific filters is machine learning filters.

17. The system of claim 1, further comprising a component that determines whether at least one IP address in the message is any one of external or internal to the recipient's system *via* a machine learning technique.

18. The system of claim 17, the component employs MX records to determine a true source of a message by way of tracing back through a received from list until an IP address is found that corresponds to a fully qualified domain which corresponds to an entry in the domain's MX record; and determines whether the IP address is external or internal by performing at least one of the following:

concluding that the IP address is in a form characteristic to internal IP addresses; and

performing at least one of an IP address lookup and a reverse IP address lookup to ascertain whether the IP address correlates with a sender's domain name.

19. The system of claim 17, the component determines whether the IP address is external or internal comprises at least one of the following:

collecting user feedback related to user classification of messages as spam or good;

examining messages classified as good by a user to learn which servers are internal; and

finding a worst-scoring IP address in a message.

20. A system that facilitates spam detection comprising:

a component that receives an item and extracts a set of features associated with an origination of a message or part thereof and/or information that enables an intended recipient to contact or respond to the message;

at least one filter that is used when one of the IP address of the message or at least some part of at least one of the URLs in the message is known.

21. The system of claim 20, the at least one filter is trained on one of known IP addresses or known URLs together with text-based features.

22. The system of claim 20, further comprising at least one other filter that is used to examine text-based features in the message.

23. A machine learning method that optimizes an objective function of the form

OBJECTIVE(MAXSCORE(m1), MAXSCORE(m2), …,
MAXSCORE(mk), w1…wn) where MAXSCORE(mk) =
MAX(SCORE(IPk,1), SCORE(IPk,2), ..., SCORE(IPk,kl))

where mk = messages;

IPk,i represents the presence of some property(s) of mk; and

SCORE(IPk,i) = the sum of the weights of the features of IPk,i.

24. The machine learning method of claim 23, the objective function depends in part on whether the messages are properly categorized as any one of spam or good.

25. The machine learning method of claim 23, further comprises learning the weights for each feature in turn.

26. The machine learning method of claim 25, learning the weight for a given feature comprises sorting training instances comprising a property, the property comprising a feature in order by the weight at which the score for that message varies with the weight for that feature.

27. The machine learning method of claim 26, the training instances comprise electronic messages.

28. The machine learning method of claim 23, the messages are training instances and the property and the properties comprise one or more IP addresses that the message originated from and any URLs in the message.

29. The machine learning method of claim 23, learning is performed using an approximation $MAX(a_1, a_2, \ldots, a_n)$ is approximately equal to $SUM(a_1^x, a_2^x, \ldots, a_n^x)^{(1/x)}$.

30. The machine learning method of claim 29, the objective function depends in part on whether the messages are properly categorized as spam or good.

31. A method that facilitates spam detection comprising:

providing a plurality of training data;

extracting a plurality of feature types from the training data, the feature types comprising at least one IP address, at least one URL and text-based features; and

training a plurality of feature-specific filters for the respective feature in an independent manner so that a first feature does not unduly influence a message score over a second feature type when determining whether a message is spam.

32. The method of claim 31, the plurality of training data comprises messages.

33. The method of claim 31, the plurality of feature-specific filters comprises at least two of the following:

a known IP address filter;

an unknown IP address filter;

a known URL filter;

an unknown URL filter; and

a text-based filter.

34. The method of claim 33, the known IP address filter is trained using 32 bits of IP addresses.

35. The method of claim 33, the unknown IP address filter is trained using some number of bits of IP addresses less than 32 bits.

36. The method of claim 33, the unknown IP address filter is trained using other messages comprising unknown IP addresses.

37. The method of claim 33, the text-based filter is trained using words, phrases, character runs, character strings, and any other relevant non-IP address or non-URL data in the message.

38. The method of claim 33, employing at least one of the known IP address filter, the unknown IP address filter, the known URL filter, and the unknown URL filter together with the text-based filter to more accurately determine whether a new message is spam.

39. The method of claim 33, further comprising employing at least one of the feature-specific filters in connection with determining whether a new message is

spam, such that the feature-specific filter is selected based in part on most relevant feature data observed in the new message.

40. The method of claim 33, the URL filter is trained on URL data comprising a fully qualified domain name and subdomains of the fully qualified domain name.

41. The method of claim 31, further comprising combining message scores generated from at least two filters used to scan a new message to generate a total score that facilitates determining whether the message is spam.

42. The method of claim 41, combining message scores comprises at least one of the following:

multiplying the scores;

adding the scores; and

training a new model to combine the scores.

43. The method of claim 33 combined with a feedback loop mechanism whereby users provide their feedback regarding incoming messages by submitting message classifications to fine tune the one or more feature-specific filters.

44. The method of claim 31, further comprising quarantining messages that satisfy at least one criterion for a period of time until additional information about the message can be collected to update one or more feature-specific filters to facilitate determining whether the messages are spam.

45. A data packet adapted to be transmitted between two or more computer processes facilitating improved detection of spam, the data packet comprising: information associated with training a plurality of feature-specific filters in an independent manner to mitigate undue influence between features and employing

at least one feature specific filter comprising an IP address filter or a URL filter to determine whether a message is spam.

46. A computer readable medium having stored thereon the system of claim 1.

47. A spam detection system comprising a plurality of filters comprising at least one filter that is trained by using different smoothing for different spam features.

48. The system of claim 47, the feature is one of the following: an IP address or a portion thereof or a URL or a portion thereof.

49. The system of claim 48, the at least one filter is trained by using different smoothing for different portions of at least one of an IP address or a URL.

50. A method that facilitates spam detection comprising:

extracting data from a plurality of messages;

training at least one machine learning filter using at least a subset of the data, the training comprising employing a first smoothing for at least one of IP address or URL features and at least a second smoothing for other non-IP address or non-URL features.

51. The method of claim 50, the smoothing differs in at least one of the following aspects:

the first smoothing comprises a different variance compared to the second smoothing with respect to a maximum entropy model; and

the first smoothing comprises a different c value for an SVM model.